DiCE: Dichoptic Contrast Enhancement for VR and Stereo Displays

FANGCHENG ZHONG, Dept. of Computer Science and Technology, University of Cambridge
GEORGE ALEX KOULIERIS, Durham University & Université Côte d'Azur, Inria
GEORGE DRETTAKIS, Université Côte d'Azur, Inria
MARTIN S. BANKS, UC Berkeley
MATHIEU CHAMBE, ENS Rennes
FRÉDO DURAND, MIT CSAIL
RAFAŁ K. MANTIUK, Dept. of Computer Science and Technology, University of Cambridge

In stereoscopic displays, such as those used in VR/AR headsets, our eyes are presented with two different views. The disparity between the views is typically used to convey depth cues, but it could be also used to enhance image appearance. We devise a novel technique that takes advantage of binocular fusion to boost perceived local contrast and visual quality of images. Since the technique is based on fixed tone curves, it has negligible computational cost and it is well suited for real-time applications, such as VR rendering. To control the trade-off between contrast gain and binocular rivalry, we conduct a series of experiments to explain the factors that dominate rivalry perception in a dichoptic presentation where two images of different contrasts are displayed. With this new finding, we can effectively enhance contrast and control rivalry in mono- and stereoscopic images, and in VR rendering, as confirmed in validation experiments.

CCS Concepts: • **Computing methodologies** → **Perception**; **Virtual reality**; *Rendering*; *Graphics systems and interfaces*;

Additional Key Words and Phrases: Visual Perception, Head Mounted Displays, Stereo, Binocular Vision

ACM Reference format:
Fangcheng Zhong, George Alex Koulieris, George Drettakis, Martin S. Banks, Mathieu Chambe, Frédo Durand, and Rafał K. Mantiuk. 2019. DiCE: Dichoptic Contrast Enhancement for VR and Stereo Displays. *ACM Trans. Graph.* 38, 6, Article 211 (November 2019), 13 pages.
https://doi.org/10.1145/3355089.3356552

1 INTRODUCTION

Image contrast is a critical factor of perceived image quality. Images with higher contrast are usually perceived as more realistic and 3-dimensional [Vangorp et al. 2014]. Bright, high-dynamic-range displays achieve high contrast, but may introduce flicker in low-persistence VR/AR displays and consume more power. Local tone-mapping operators can be effective at enhancing local contrast, but may lead to unnatural looking images and artifacts in video [Eilertsen et al. 2013]. They are also computationally expensive, making their use prohibitive especially in time-critical VR/AR applications, in which every GPU cycle matters and dropping frames is not an option. In this paper, we capitalize on the human visual system's binocular fusion mechanisms to enhance contrast and visual quality of images shown on binocular displays, such as those used in VR/AR headsets.

We exploit an inherent property of the binocular fusion mechanism to enhance perceived contrast. We introduce a method that

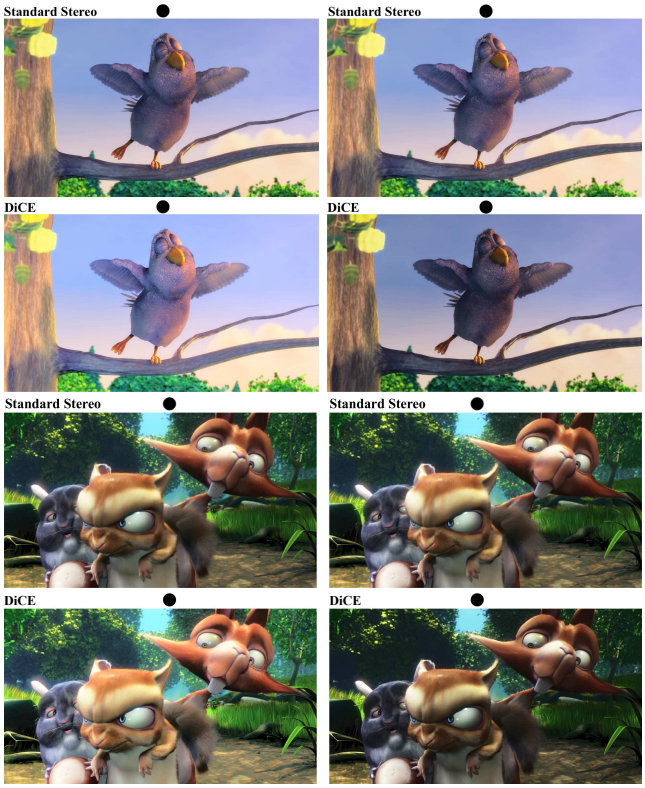

Fig. 1. Comparison of standard stereo images and the images with enhanced perceived contrast using our DiCE method. They can be cross-fused with the assistance of the dots above the images. Notice the enhanced contrast in the shadows and highlights of the scene. The stereo images are from *Big Buck Bunny* by Blender Foundation.

Authors' addresses: Fangcheng Zhong, Dept. of Computer Science and Technology, University of Cambridge; George Alex Koulieris, Durham University & Université Côte d'Azur, Inria; George Drettakis, Université Côte d'Azur, Inria; Martin S. Banks, UC Berkeley; Mathieu Chambe, ENS Rennes; Frédo Durand, MIT CSAIL; Rafał K. Mantiuk, Dept. of Computer Science and Technology, University of Cambridge.

0730-0301/2019/11-ART211 $15.00
https://doi.org/10.1145/3355089.3356552

takes advantage of this property by selectively using lower or higher tone curve slopes to improve image contrast. However, this can cause *binocular rivalry*: an unstable percept that switches between the image of one or the other eye. We conduct an experiment to establish the main factors causing this rivalry and to reliably tune the parameters of our method. From the results we find that the ratio of contrasts presented to both eyes is the main factor causing rivalry. This allows us to tune our method to greatly enhance contrast while reducing rivalry. Since the dominant cause of rivalry is mostly independent from image content, our method can be implemented as a fixed set of tone curves, which have negligible computational cost and can be directly used in real-time VR rendering. We evaluated our method by comparing to previous work, and our results clearly show that our solution is more successful at enhancing contrast and at the same time much more efficient. We also performed an evaluation in a VR setup where users indicate that our approach clearly improves contrast and depth compared to the baseline.

We make three primary contributions:

- We propose a binocular contrast enhancement method that generates a different tone curve for each eye with the intent to maximize perceived contrast while controlling the level of rivalry (Section 4).
- We experimentally establish the main factor that causes rivalry in enhanced images (Section 5). This lets us find the best parameters for our tone curve generation method (Section 6).
- We demonstrate the strengths and shortcomings of our and existing dichoptic presentation techniques in an evaluation experiment (Section 7).

Our methodology and results suggest that rendering for the binocular domain is both a computationally cheap and effective means to increase contrast in binocular displays.

2 RELATED WORK

The proposed method is a post-processing technique that can be combined with any existing tone mapping or applied directly to monoscopic or stereoscopic images. Because it shares goals and techniques of tone mapping, we first briefly review tone-mapping methods. Then we discuss previous methods that rely on binocular fusion.

2.1 Tone Mapping and Contrast Enhancement

Tone mapping is an image-processing operation performed to convert an image from a scene-referred color space into a display-referred color space. Tone mapping spans a range of techniques that can vary in their goals. Some techniques simulate specific phenomena of the visual system (glare, night vision). Others attempt to achieve the best subjective quality (color grading, enhancement) or possibly faithful reproduction of image appearance [Eilertsen et al. 2017]. Because scene-referred colours often exceed the dynamic range of the target display, a common goal of all tone-mapping methods is the reduction of dynamic range.

One of the most common techniques used in tone mapping is a global tone curve: a monotonic function that maps input color/luminance values to the displayed color/luminance values. Such a curve can be fixed and, for example, can mimic the response of a photographic film [Reinhard et al. 2002], or can adapt to image content [Ward-Larson et al. 1997] and a display [Mantiuk et al. 2008]. A tone curve is typically designed to enhance contrast in visually relevant parts of the scene and compress or clip contrast in less relevant parts, which are dark or noisy [Eilertsen et al. 2015], or contain bright highlights or light sources that cannot be easily reproduced on a display.

To revert the loss of small contrast details caused by compressive tone curves, many tone-mapping techniques involve local contrast enhancement. Such enhancement could be achieved by unsharp masking combined with edge-stopping filters [Durand and Dorsey 2002; Eilertsen et al. 2015], which can avoid ringing or halo artifacts. Stronger enhancement could be achieved by operating in the gradient domain [Farbman et al. 2008]. This, however, requires computationally expensive optimization. Contrast at multiple scales can be more efficiently edited using local Laplacian pyramids [Paris et al. 2011]. The main drawback of all these enhancement techniques is that they introduce a substantial computational overhead, which is unacceptable in real-time applications. Our technique replaces computationally expensive local contrast enhancement with fixed tone curves, which have negligible computational cost.

2.2 Exploiting the Binocular Domain

Binocular fusion was exploited before in a number of tone-mapping methods for binocular displays [Yang et al. 2012; Zhang et al. 2019, 2018]. We will refer to these methods as BTMO (Binocular Tone Mapping Operators). The goal of these techniques is to produce two tone-mapped images that are maximally different, yet comfortable to fuse. This is achieved by adjusting the parameters of an existing [Yang et al. 2012] or newly proposed tone-mapping operator [Zhang et al. 2018] in an optimization loop. The loss function is designed to maximize the difference between left- and right-eye images, leading to "richer" fused images. To ensure acceptable levels of rivalry, a *binocular viewing comfort predictor* is used to reject image pairs that are deemed too rivalrous. Neural networks can also be leveraged to generate tone-mapped images without assumptions about monocular tone-operators. In concurrent work, *Deep binocular tone mapping* [Zhang et al. 2019] employs CNNs to generate an end-to-end binocular tone mapping operator that outputs the desired LDR pair from an HDR image. Similar to previous BTMO techniques, the loss function is designed to optimize the visual content distribution to maximize the perception of local detail and global contrast, while maintaining visual comfort. Real-time computation can be achieved with a GPU acceleration.

In contrast to BTMO techniques, our method explicitly enhances contrast based on psychophysical models and findings, rather than making images different. In Section 7, we demonstrate that this leads to much more consistent and predictable enhancement. Instead of a complex viewing comfort predictor, which combines multiple heuristics, we find a simple yet effective rivalry indicator based on new experimental findings. Our technique does not restrict the choice of tone-mapping operator and can be used with stereoscopic content. But most importantly, our technique has negligible computational cost compared to the BTMO methods, and thus can process

an image pair in milliseconds rather than seconds without relying on GPUs.

3 BACKGROUND

We first provide definitions of physical quantities and their perceptual correlates central to understanding the principles of human vision that govern the perceptual phenomena we exploit. Then, we discuss the models and relations that explain the fusion of dichoptic stimuli.

3.1 Physical quantities and their perceptual correlates

One of the fundamental findings in psychophysics are those about non-linear and complex relations between the physical stimuli, and the perceptual sensations they elicit. Our method relies on matching perceived rather than physical quantities, so it is important to explain such relationships.

The photometric measure of the luminous intensity per unit area of light, measured in cd/m^2 (nits) is *luminance*. The difference in luminance that makes an object or its representation on a display distinguishable from its surroundings is *contrast*. Contrast is typically expressed using Michelson's formula (refer to Figure 5), but there are alternative formulas: Weber, logarithmic, and energy-based RMS contrast. The relationship between physical contrast and perceived contrast magnitude can be explained by a power function with an exponent ranging from 0.45–0.55 [Cannon 1985]. The contrast between the maximum and minimum luminance that can be produced on an electronic display is its dynamic range. Although it is typically reported as a contrast ratio L_{max}/L_{min}, a more perceptually uniform measure of dynamic range is given by the difference of log-luminance, $\log_{10}(L_{max}) - \log_{10}(L_{min})$.

3.2 Perception in Dichoptic Presentation

In a binocular display, *dichoptic* presentation is the presentation of different images to the two eyes and *dioptic* is the presentation of identical images to the two eyes. If the dichoptically viewed images are synthesized/photographed from two offset viewpoints at a distance approximately equal to the human interpupillary distance, they contain image disparities that elicit the illusion of depth by exploiting binocular vision. This is a *stereoscopic* image pair and always requires dichoptic presentation. Dioptic presentation cannot ellicit the illusion of depth from disparities — as images for left and right eye are identical — and thus can only show *monoscopic* images. To avoid confusion, we will refer to images without dichoptic enhancement as *standard* in the manuscript, regardless of whether these are monoscopic or stereoscopic images.

When the dichoptic stimuli are too dissimilar to be fused into one stable percept, the viewer experiences *binocular rivalry*. Binocular rivalry refers to a state of competition between the eyes, with one eye inhibiting the perception of the image in the other eye causing alternation between perceived images [Blake 1989]. Rivalry is caused primarily by geometric differences in the two eyes' images. A special case is *luster*, which occurs when luminance or contrast differences exist in corresponding image areas. It creates a shiny appearance in such areas.

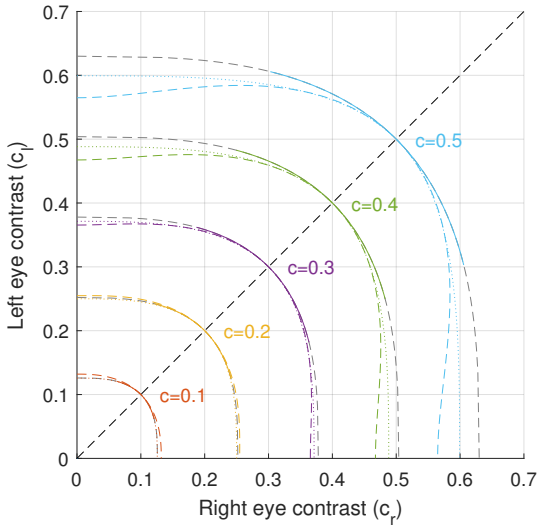

Fig. 2. For each level of standard (dioptic) contrast ($c = 0.1 \,..\, 0.5$), the color lines show the combination of the left and right eye contrast (dichoptic contrast) that produces the match. The color lines represent the standard and the axis the test conditions in the contrast matching experiment. The lines are plotted according to the contrast matching model from Equation 2 and assuming $\beta = 3$. The black-dashed line represents standard contrast. The grey-dashed lines illustrate the range of contrast combinations that result in an unstable percept and rivalry. The color dashed lines illustrate the same relation but according to the late summation model (Equation 3), and the dotted colour lines show the relation in terms of logarithmic contrast (Equation 4).

Fusion of luminance. When a uniform patch of luminance L_l is shown to the left eye, and a patch of luminance L_r to the right eye, the fused patch can be matched to the luminance that is the (weighted) average of those:

$$L_{\text{fused}} = a\,L_l + (1-a)\,L_r\,, \tag{1}$$

where L_{fused} is the matching luminance (presented to both eyes) and a compensates for the dominant eye [Levelt 1965] ($a \approx 0.5$).

Fusion of contrast. Legge and Rubin [1981] investigated perceived contrast when two stimuli of the same spatial configuration but different contrasts are presented to the two eyes. Two stimuli were presented: The *standard* in which the same contrast is presented to the two eyes and the *test* in which a different contrast is presented to each eye. The subject adjusted the contrast of the test in one eye to create the same perceived contrast for the standard and test. They found that a generalized mean best describes their data. If we present contrast c_L to the left eye and contrast c_R to the right eye, the magnitude of the perceived, matched standard/dioptic contrast

c_m is:

$$c_m = \left(\frac{c_L^\beta + c_R^\beta}{2} \right)^{\frac{1}{\beta}} . \quad (2)$$

β tends to be close to 3. It is the same across spatial frequencies and increases slightly with contrast. The matching contrast obtained by the above formula is illustrated as color curves in Figure 2. The curves show that the fused contrast is dominated by the eye with the stronger contrast, in a manner that is close to the winner-take-all strategy.

Kingdom and Libenson [2015] further show that the contrast fusion can be explained by the late summation model in which the signals from both eyes contribute to the response, R, of a contrast transducer function:

$$R(c_L, c_R) = \frac{c_L^p + c_R^p}{z + c_L^q + c_R^q} , \quad (3)$$

where z, p, and q are the parameters controlling the shape of the contrast transducer [Legge and Foley 1980]. Curves of matching contrast resulting from the late summation model are shown as dashed color curves in Figure 2. Because both models are comparable in the range where inter-ocular contrast differences are small (and the rivalry is low), we will rely on the simpler form in Equation 2 in the further analysis.

3.3 Partial overlap HMDs

Binocular human vision is achieved via two monocular visual fields of around 160° of horizontal visual angle each; their total horizontal field of view is approximately 200°. The combined FoV consists of three regions: an overlapping 120° central binocular region where stereopsis is achieved and two flanking monocular regions of approx. 40° each [Palmer 1999].

Older HMDs employed a full overlap design, in which both eyes saw the same part of the scene. This resulted in smaller field-of-views as the optical design was limited by the human binocular region. Modern commercial HMDs have a partial overlap design, mimicking the human visual system. This allows for physically smaller displays while both increasing the FoV and thus immersion, and supporting wider aspect ratios [Fuchs 2017]. In such HMDs, binocular overlap refers to the visible overlapping portion between the two eyes (see Fig. 3) in the headset and describes how much of the virtual scene can be seen by both eyes, which is crucial for depth perception. In partial overlap binocular displays only a central region of the scene is shown to both eyes, and areas to either side are seen by only one eye. This often creates interocular differences in the monocular regions [Patterson et al. 2007] which often induce a perceptual effect known as luning which is the subjective darkening in the monocular regions or, for other users, it is experienced as a visual fragmentation of the field-of-view into three distinct regions (left, middle, right) [Klymenko et al. 1994]. In modern headsets this can be alleviated by post-processing the binocularly overlapping regions.

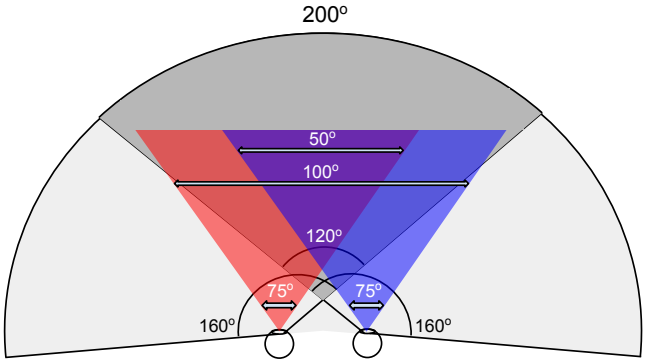

Fig. 3. Arcs denote angles for viewing in the real world: each eye sees a field of view of about 160°. This results to a 200° combined horizontal field of view, 120° of which are overlapping and thus binocular processing or stereopsis is possible. Two-headed arrows denote angles in a modern VR headset: each eye sees a horizontal FoV of about 75°, leading to a 100° combined FoV, only 50° of which are overlapping and available for binocular processing or stereopsis.

4 DICHOPTIC CONTRAST ENHANCEMENT

In this section we explain how the contrast of images seen binocularly can be enhanced beyond what can be reproduced on a typical display significantly improving image quality and realism in VR headsets and stereo displays. Our method was inspired by the observation of Legge and Rubin that the fused contrast is dominated by the image of higher contrast (Equation 2). We take advantage of stereoscopic displays, which can present a different image to each eye and therefore offer a separate dynamic range budget for the left and right eye. This lets us selectively use lower or higher tone curve slopes to improve image contrast. When binocularly fused, the images convey more fine detail in the shadows and highlights compared to standard tone-mapped images.

4.1 Tone curves and contrast enhancement

We define a tone curve as a function mapping the logarithmic luminance (base-10 logarithm) of the input image to the physical logarithmic luminance of the display device, as shown in Figure 4. Representing luminance in the logarithmic domain makes it more perceptually uniform (see Sec 2.4 in [Mantiuk et al. 2015]) but also has the property that the slope of the tone curve in the log-log domain modulates contrast of the corresponding tonal range. Altering the slope corresponds to multiplying log-luminance values: i.e., raising linear luminance values to a power (commonly known as gamma).

A well-selected tone curve can achieve high contrast in any relevant tonal range while mapping all pixel values to the available dynamic range. Assigning a steeper slope in one part of the tone curve boosts contrast in that range, however, a larger proportion of the output dynamic range budget is spent, necessitating contrast compression in another part of the input range. The output log-luminance is restricted by the peak luminance of the display (d_{max}) and its black level (d_{min}).

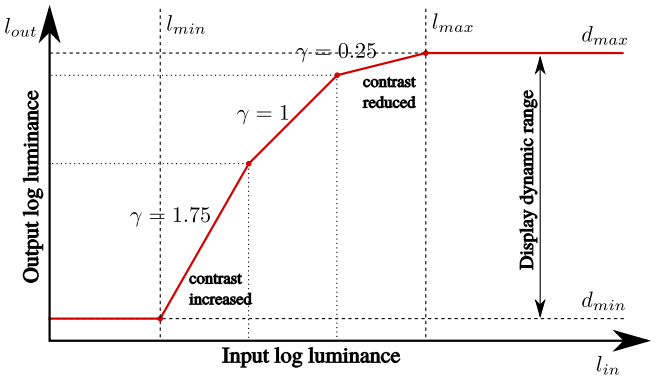

Fig. 4. An example tone curve mapping input image log luminance to output image log luminance. The slope of the tone curve corresponds to the reduction or increase in contrast in the given tonal range of an image.

To ensure that we can rely on the contrast fusion rule when manipulating tone curves, we need to address the discrepancy in contrast units. The contrast fusion rule in Equation 2 is defined in terms of Michelson contrast, which we denote as c. The slope of the tone curve directly alters logarithmic contrast, which we denote as g. Logarithmic contrast is defined as a half of the logarithm of the luminance ratio, as illustrated in Figure 5. Logarithmic contrast is not equivalent to Michelson contrast. However, for small and medium contrasts ($c < 0.5$) that dominate natural or computer-generated imagery, both contrast measures are linearly related, as shown in Figure 5. Thus, the contrast fusion can be expressed in terms of logarithmic contrast:

$$g_m = \left(\frac{g_L^\beta + g_R^\beta}{2} \right)^{\frac{1}{\beta}} . \tag{4}$$

The new contrast matching formula, plotted as dotted lines in Figure 2, predicts contrast match that lies between the predictions of Equations 2 and 3.

4.2 Interleaved dichoptic tone curves

Let us consider how we can design a tone curve that would maximize contrast enhancement within the given budget of the dynamic range. A simple approach would be to create two tone curves, like those in Figure 6, consisting of two piece-wise linear segments. For a given tone curve segment, the slope in one eye can be increased while reduced in the other without exceeding the dynamic range budget. If the base tone curve (black dashed line in Figure 6) has the slope s_b, we set the slope for one eye to s_l and the slope for the other eye to $s_h = 2\,s_b - s_l$ so that $s_l + s_h = 2\,s_b$. We will use indices l and h to denote low and high slope (rather than left and right eye) as the slopes will be assigned interchangeably to each eye for each segment of the tone curve. From Equation 2, we can find that the gain in fused contrast for the original contrast g is:

$$\Gamma = \frac{1}{g\,s_b} \left(\frac{(g\,s_l)^\beta + (g\,s_h)^\beta}{2} \right)^{\frac{1}{\beta}} = \frac{1}{s_b} \left(\frac{s_l^\beta + s_h^\beta}{2} \right)^{\frac{1}{\beta}} . \tag{5}$$

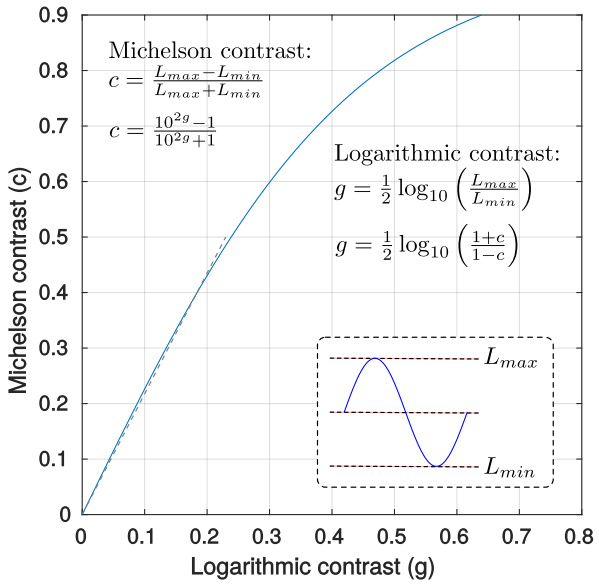

Fig. 5. The relation between logarithmic and Michelson contrast.

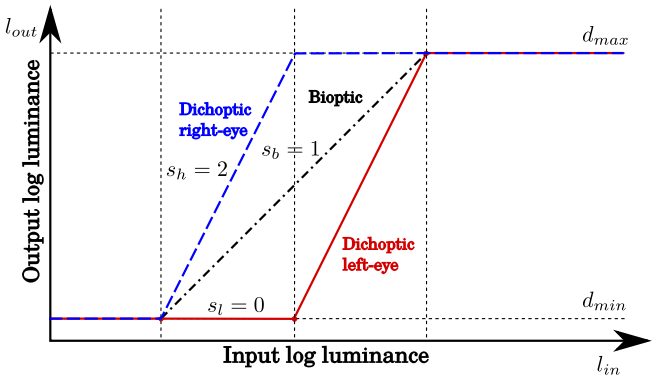

Fig. 6. When a separate tone curve is used for each eye (dichoptic presentation) the slope of one of the curves can be up to twice as high as that for a standard dioptic presentation. The perceived contrast for the dichopic images will be 10%–50% higher (see Figure 7). However, such strong separation of the tone curves will result in an image that is very uncomfortable to view.

The gain as the function of the slope on the left and right eye is plotted in Figure 7. The curves clearly show that the gain in perceived contrast is greatest when the slope is maximized in one eye and minimized in another. However, such a large luminance and contrast difference could result in strong binocular rivalry.

To reduce the luminance difference and thus the potential cause of rivalry, we want the left- and right-eye tone curves to be more similar to each other. This can be achieved with an interleaved tone curve with a higher number of piece-wise linear segments, such as the one in Figure 8. It should be noted that increasing the number of segments does not affect the slopes of the curves in the left and right

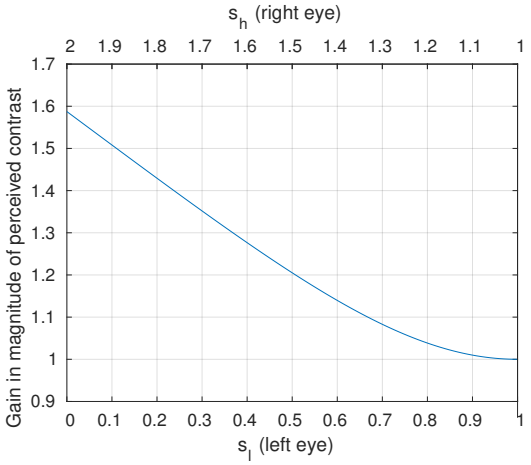

Fig. 7. The gain in contrast due to fusing left and right eye images which are processed by the tone curves with the slopes s_l and s_h (x-axis). As the tone curve slope is reduced on the left eye (s_l), it is increased on the right eye (s_h). Such a change in slope does not reduce the dynamic range budget allocated to both eyes, but it boosts fused contrast.

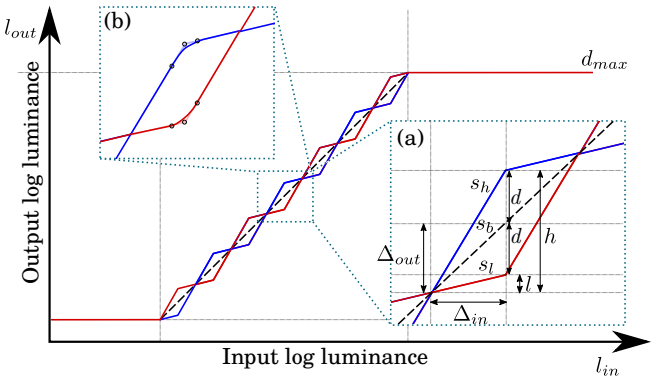

Fig. 8. Binocular tone curves may introduce less rivalry if they are constructed so that resulting luminance values in each eye are possibly similar. The interleaved low- and high-slope segments could be used to produce such curves. Inset (a) shows the notation we use. We denote the lower slope as $s_l = l/\Delta_{in}$ and the higher slope as $s_h = h/\Delta_{in}$. We also denote the number of linear segments in the tone curve as N. For example, the segment that spans Δ_{in} is what we mean by one segment. Inset (b) shows smoothing using Bezier curves. The black circles denote the control points.

eyes and therefore does not affect contrast enhancement. However, the number of segments restricts the highest contrast that can be manipulated by the tone curve: if the contrast between two pixels is large enough to span two segments of the tone curve (i.e. be larger than Δ_{in}), it is not going to be enhanced (or reduced) as intended. Finding the right number of segments and their slopes is a challenging problem and we address this problem in a series of experiments in Section 5. But first, we explain why we need to ensure smoothness of the tone curves.

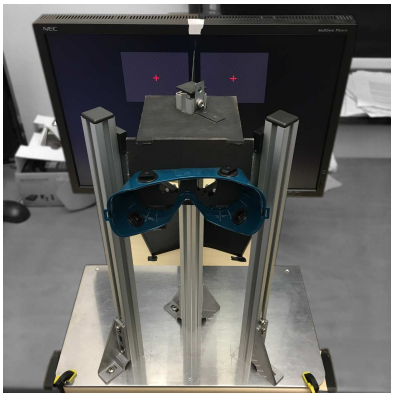

Fig. 9. LCD display with a stereoscope used in the experiments.

4.3 Smooth tone curves

In preliminary experiments, we observed that the piece-wise linear interleaved tone curves may result in banding artifacts when an image contains large areas with smooth gradients. These are caused by the C^1 discontinuities in our tone curves, which translate to similar discontinuities in the resulting image. The visual system is very sensitive to such discontinuities, which are interpreted as spurious contours [Kingdom and Moulden 1988]. This problem can be easily addressed by replacing the small intervals containing discontinuities in the piece-wise linear curve with a cubic Bezier curve. We set the size of the interval to be 0.1 log_{10} units. The three control points of this Bezier curve are the two end points on the interval and the slope-transition point, as shown in 8-(b). This ensures that our tone curves are C^1 continuous in the entire domain.

5 EXPERIMENT 1: THE PREDICTOR OF RIVALRY

The interleaved dichoptic tone curves are controlled by two parameters: the number of segments and the slope of the interleaved tone curves. To determine the optimal choice of these parameters that would produce the strongest enhancement and acceptable level of rivalry, we conducted a perceptual experiment. The experiment was intended to test two hypotheses, each proposing a different indicator of binocular rivalry:

Hypothesis 1. If rivalry is induced by the luminance difference between the left and right eyes, a good predictor would be the maximum log-luminance difference, or $h-l$ using the notation from Figure 8. Note that $h - l = (s_h - s_l)\,\Delta_{in}$.

Hypothesis 2. Rivalry may also be caused by the contrast difference between the left and right eyes. A good predictor in this case would be the ratio of contrasts presented to the two eyes $s_l/s_h = l/h$.

Apparatus and Participants. The experiment was performed on a 24-inch NEC PA241W colorimetrically calibrated display with an attached stereoscope in a dark room (Figure 9). The optical path to the display was 36cm (2.77D). Eight volunteers participated (mean age 27.3, SD 4.2 years). Before the experiment, each participant

read and signed the consent form. We also demonstrated to each participant what rivalrous and non-rivalrous stimuli look like.

Stimuli and Procedure. We selected 16 HDR images, which were tone mapped based on the smooth inter-leaved tone curves with N equal segments as explained in Section 4. The end-points of the tone curve were set to be at the 1st and 99th percentiles of image luminance. The dynamic range of the target display was 2.7 log-10 units (500:1 contrast).

The participants were asked to adjust the deviation d (shown in Figure 8) from the straight tone curve so that "the image looks sharp and comfortable to view" (exact wording on the briefing form). The critical values of d were measured using the method-of-adjustment procedure with three repetitions per image. Then, the two proposed predictors were computed accordingly as:

$$h - l = 2d \tag{6}$$

$$\frac{l}{h} = \frac{\Delta_{out} - d}{\Delta_{out} + d} \tag{7}$$

The experiment consisted of six sessions. The same HDR images were used in all of them. Four of the sessions had $N = 2, 4, 10$, and 20 interleaved segments spanning the entire dynamic range of the display, 2.7 log-10 units. The two remaining sessions had $N = 10$ segments spanning half of the display's dynamic range, 1.35 log-10 units, so that one session spanned the darker half and one the brighter half of the dynamic range. Example of images rendered with different number of segments and slopes are shown in Figure 10. The order of sessions and images was randomized.

5.1 Results

The plots for the two proposed predictors and for eight participants are shown in Figure 11. It is evident that the ratio of contrast l/h is a much more consistent predictor than the log-luminance difference across different test conditions (number of segments, output display dynamic range). This was further tested in a leave-one-out cross-validation, where we used the 7 out of 8 of the measured images to calculate a fixed value of the predictor, which was then used to predict the s_l values of the remaining images. The procedure was repeated eight times. The prediction error was computed as RMSE between the true and predicted s_l and is shown in Table 1. The results suggest that the ratio of contrast l/h is indeed the better predictor for s_l.

Discussion. The results demonstrate that the magnitude of rivalry is determined by the contrast difference between the eyes (Hypothesis 2) rather than by the luminance difference (Hypothesis 1). This finding confirms the importance of contrast in visual processing [Kingdom and Libenson 2015]. There is ample evidence suggesting that low-level visual mechanisFms attempt to preserve contrast but they do not encode information about absolute luminance. For example, Weber's law states that we are sensitive to ratios (contrast) rather than absolute levels. Contrast constancy preserves the appearance of supra-threshold contrast across spatial frequency and to some extent across luminance range [Georgeson and Sullivan 1975; Kulikowski 1976]. Furthermore, light-/dark-adaptation is attributed in large extend to the retina (photoreceptors and bipolar cells) [Dunn and Rieke 2008] and can be controlled individually per eye. This means that a per-eye luminance difference can be partially compensated by the adaptation mechanism. Therefore, it is not surprising that conflicting contrast signals evoke more rivalry than conflicting luminance signals. This finding also shows that some degree of rivalry is unavoidable as we need to introduce contrast differences for contrast enhancement. However, many observers reported that they can adapt to a moderate level of rivalry a few seconds after switching from standard to dichoptic presentation.

Table 1. Each row represents the prediction errors (RMSE) for each participant using the corresponding predictors.

Participant	Log-luminance difference	Ratio of contrast
1	0.4182	0.0870
2	0.3793	0.0971
3	0.2880	0.0795
4	0.2062	0.0576
5	0.3811	0.0895
6	0.5217	0.1541
7	0.2824	0.0892
8	0.3148	0.0981

It should be also noted that the ratio of contrast l/h as a predictor of rivalry is independent of image content. As shown in Figure 11, we cannot observe a pattern for images that would be consistent across the participants. The differences in the means between observers are also small given the within-observer variance. Therefore, the high variance is likely to be due to the measurement noise, rather than systematic effects. To further validate it, we also conducted a similar experiment for sinusoidal gratings instead of complex images to better isolate the effect of contrast, luminance, and spatial frequency. We did not find a significant effect of any of these factors. We include the details of this experiment in the supplemental material (Experiment 1a).

5.2 Rivalry due to luminance difference

Although the contrast seems to be the dominant factor in dichoptic rivalry, we cannot fully discount the effect of luminance. If we did so, we would need to assume that two images of the same contrast but very different luminance are always comfortable to fuse. To determine maximum luminance difference that can be regarded as acceptable, we conducted one additional experiment (Experiment 1b) using the same protocol as in Experiment 1 and described in detail in the supplementary material. The results indicated that most observers can tolerate the luminance difference $(h - l)$ up to 0.66 log-10 units (50th percentile). We will use this result to determine the best number of segments in Section 6.1.

6 IMPLEMENTATION

Experiment 1 demonstrated that binocular rivalry is mostly induced by the contrast difference between the eyes. The variance in the perceived rivalry between the images is relatively small, therefore, we can make our enhancement method independent of image content. Our interleaved tone curves can be precomputed, and applied to an image after tone mapping (but before display coding). This is

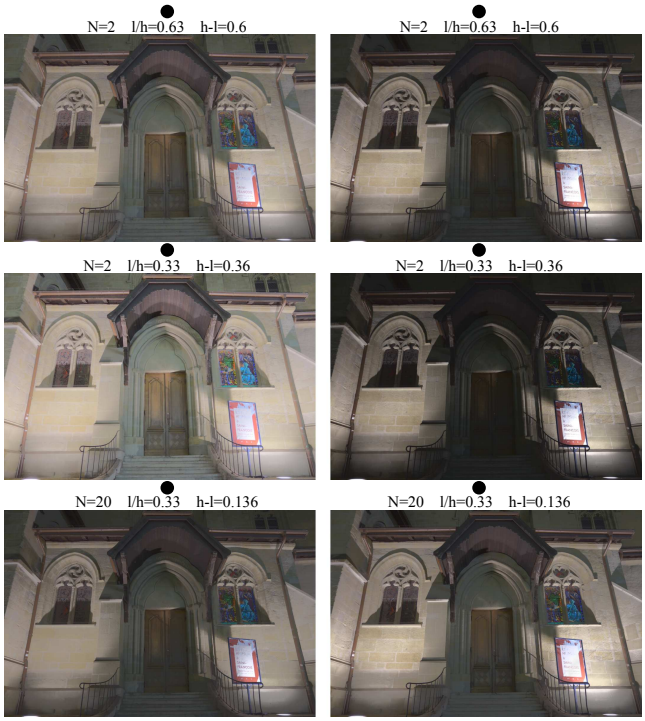

Fig. 10. Examples of DiCE-enhanced monoscopic images from Experiment 1, with different strength of enhancement (the enhancement is stronger at a low l/h ratio), and different number of segments of interleaved tone curves (N). The images are suitable for cross-fusion.

a significant advantage of our DiCE method, letting us use it with any existing tone-mapping operator, or directly with SDR images.

Figure 12 shows the diagram of a tone-mapping pipeline with DiCE. First, any existing tone-mapping operator can be used to reduce the dynamic range of an HDR frame and generate a display-referred frame. Alternatively, a SDR frame, decoded into a linear RGB space, can be used as input to our method. We then separate a luminance channel from CIE $u'v'$ chromacities and apply the interleaved tone curves to the luminance channel alone. The color is added back using an inverse color transformation. Finally, the colors are display encoded and stored in a raster buffer. Depending on the target display, they can be encoded into the sRGB space for SDR displays, or one of the color spaces from the ITU BT.2100 recommendation for HDR displays.

6.1 Selecting interleaved tone-curve parameters

Our experimental results indicate that l/h determines both contrast enhancement and the magnitude of rivalry. The l/h is also independent of the number of segments. Given that, we opt for the smallest number of segments for two reasons: a) wider segments let us enhance a broader range of spatial frequencies (as discussed in Section 4.2); and b) there is a smaller chance for banding artifacts in the region where the tone curve switches from low to high slope (as discussed in Section 4.3). However, small number of segments increases the maximum luminance difference, which could be another

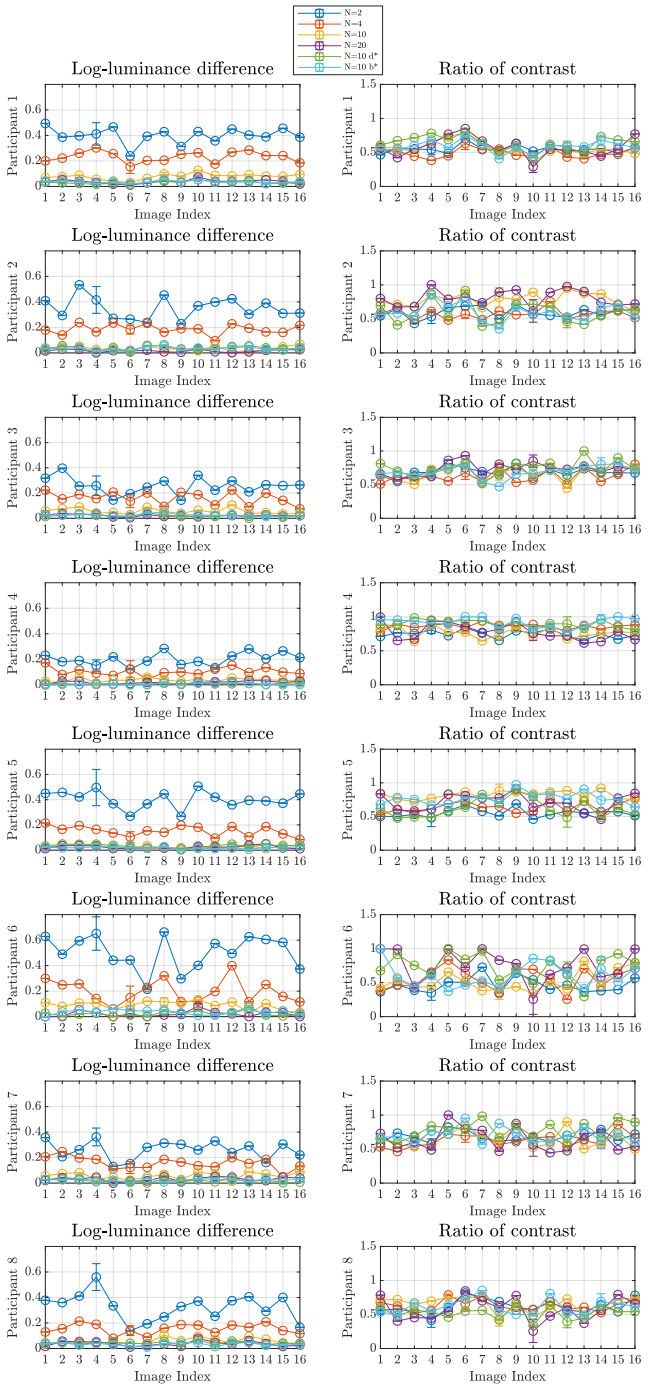

Fig. 11. The two proposed predictors of binocular rivalry (columns) collected from Experiment 1, for eight participants (rows). The colors denote different number of segments N and different output display dynamic range (d* and b* indicate half of the display's dynamic range, with d* representing the darker half and b* representing the brighter half). The error bars represent the expected value of the standard deviation for the given set of conditions. It is evident that the ratio of contrast l/h is distributed more uniformly than the log-luminance difference.

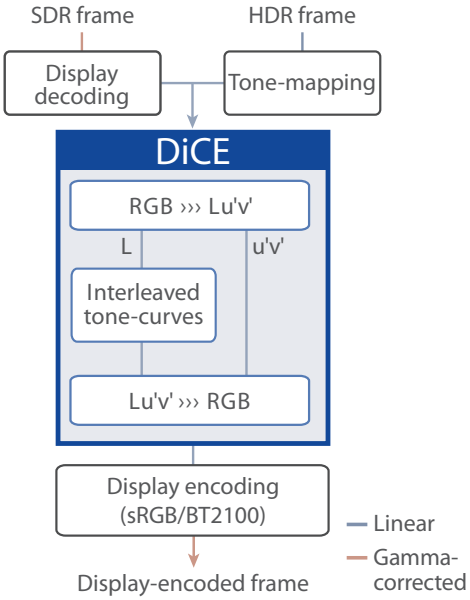

Fig. 12. DiCE as part of a tone-mapping pipeline. The dynamic range of HDR input frames (in linear RGB color space) can be reduced with any tone mapping operator. Alternatively a standard SDR frame can be used. The luminance is separated from two color-opponent channels. The per-eye interleaved tone curves are applied to the luminance channel, separately for each eye and then color is added back. Finally, the pixel values are display-encoded into SDR (sRGB) or HDR (rec.2100) display-referred space.

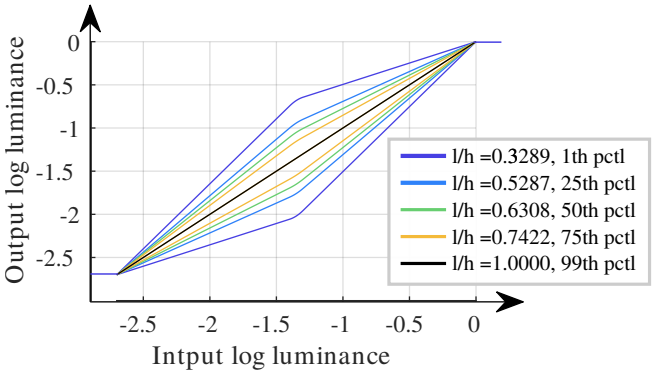

Fig. 13. The shape of the dichoptic tone-curves at different l/h ratios. The ratios were selected to represent 1st, 25th, 50th, 75th and 99th percentile of the data (across all images and observers) from Experiment 1.

cause of rivalry, as discussed in Section 5.2. Therefore, in Figure 14 we plot the maximum luminance difference $(h - l)$ as a function of the display dynamic range and the number of segments (Experiment 1b, in supplementary). The plots show that $N = 2$ is the right choice for most SDR displays up to 2.8 log-10 units of the dynamic range, including OLED displays used in HTC Vive and Oculus Rift. The number of segments, however, may need to be increased to 4 for high-contrast HDR displays.

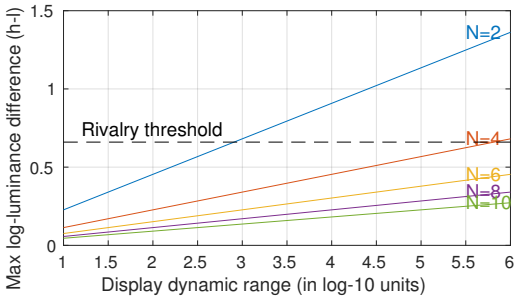

Fig. 14. The maximum log-10-luminance difference $(h - l)$ for a given display dynamic range (x-axis) and the number of segments (N, colors). The plots are drawn assuming $l/h = 0.63$ (50th percentile). The dashed line represents the rivalry threshold (50th percentile) for log-luminance difference (Experiment 1b, in supplementary). The plot shows that for most SDR displays (dynamic range <2.8 log-10 units), we do not need more than 2 segments.

Slope selection for the interleaved tone curves creates a trade-off between contrast enhancement and binocular rivalry. Figure 5 shows that contrast enhancement is maximised for small ratios l/h, but, as found in the rivalry experiment, such small ratios increase binocular rivalry. Therefore, the ratio l/h should be set as a parameter, adjusted per user, ranging from about 0.5–0.75. The family of interleaved tone curves for the range of l/h ratios and two segments is shown in Figure 13.

6.2 DiCE for Partial Overlap HMDs

For partial overlap HMDs (most commercial headsets) we cannot apply the interleaved tone curves to the entire FoV. If the monocular flanking regions (magenta and blue lobes in Fig. 3) are processed by the interleaved DiCE curves, they remain unfused and show spurious contrast modulation. This is magnified with head motion in VR, which causes contrast appearance to change in the flanked regions. To avoid this problem, we employ a piece-wise linear blending function that ensures a gradual transition between the dichoptically tone-mapped area of the image that is viewed binocularly, to the monocularly tone mapped flanking lobes. The binocular overlap area depends on the fixed headset optical setup and the eye relief, i.e., the distance of the eye from the lens, which itself depends both on how deep-set the eyes are in the face and how pronounced the brow is.

7 EVALUATION

We compare our method with the standard presentation and BTMO technique on a stereoscope in Experiment 2, and then evaluate in VR rendering in Experiment 3.

7.1 Experiment 2: Validation with Stereo Display

In this experiment we compare our technique with the standard presentation (no dichoptic enhancement) and previous work (BTMO, [Zhang et al. 2018]) on a stereo display.

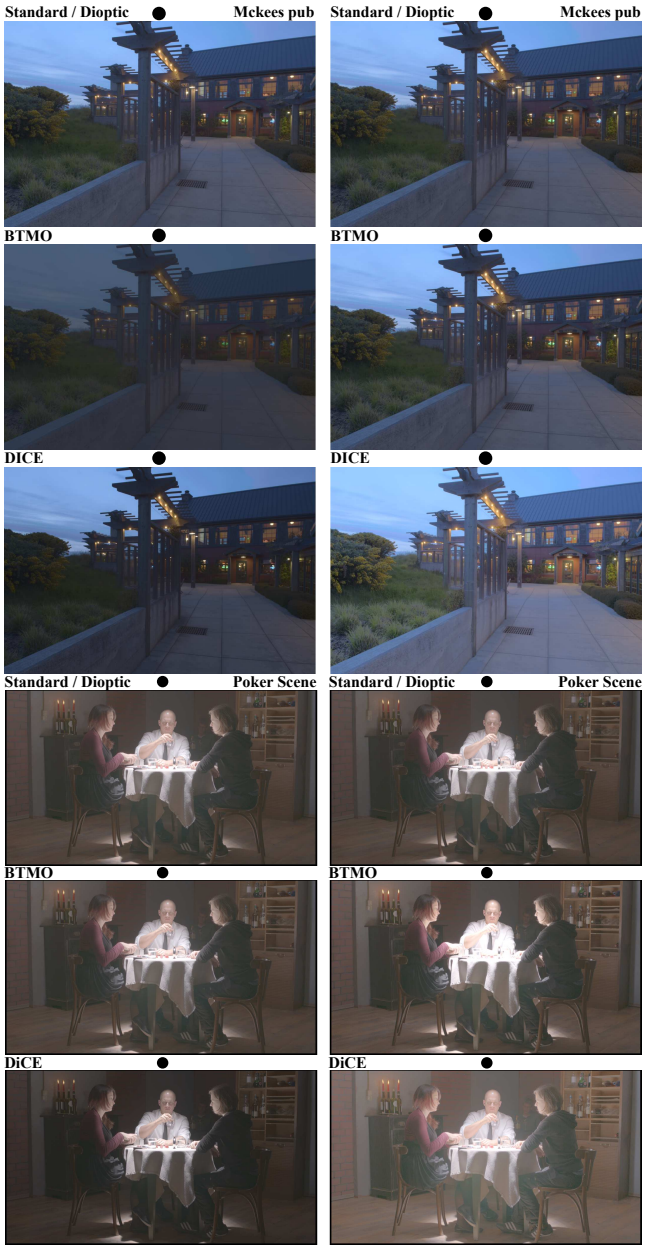

Fig. 15. Example of two images used in Experiment 2 in a format suitable for cross-fusion.

Apparatus and Participants. We used the same display and stereoscope as in the first experiment. 16 volunteers participated (5 female, mean age 26.8, SD 4.3 years).

Stimuli. 17 monoscopic images and 2 stereoscopic image were processed with our DiCE technique and the BTMO technique from [Zhang et al. 2018]. The images were kindly processed for us by the authors of the BTMO paper. It should be noted that both techniques serve a different purpose: BTMO is a tone-mapping operator that requires an HDR image as input. Our DiCE technique expects as input an image that has been already tone-mapped. Therefore, to reduce differences between the methods due to different tone-mapping operators, we used one of the images generated by BTMO as the standard/dioptic condition (no enhancement) and also as the input to our technique. When selecting an image, we chose the one from the pair (left- and right-eye image) which contained fewer under- or over-exposed pixels. We used l/h ratio of 0.63 for all DiCE-enhanced images, which was the median from Experiment 1. We selected median rather than a higher percentile as we noted that the participants are more conservative when they are asked to self-report the rivalry threshold and can tolerate higher rivalry over time, when they get used to it. Two images used in the experiment are shown in Figure 15. The remaining images can be found in the supplementary materials.

Procedure. We used a full-design pairwise comparison experiment in which all unique combinations of conditions are compared: DiCE vs. standard/dioptic, BTMO vs. standard/dioptic and DiCE vs. BTMO. The participants were asked two questions regarding each image pair that they saw: "which image has a higher contrast?" and "which image looks better?" The participants could switch between one and the other image in the pair using the arrow keys and they confirmed the image of higher contrast with the "space" key and the image they preferred with the "enter" key. Each pair was compared three times by each observer. The order of image pairs was randomized.

Data analysis. The results of the pairwise comparison experiments were scaled using publicly available software[1] under Thurstone Model V assumptions in just-objectionable differences (JODs), which quantify the relative quality differences between the techniques. A difference of 1 JOD means that 75% of the population can spot a difference between two conditions. The details of the scaling procedure can be found in [Perez-Ortiz and Mantiuk 2017]. Since JOD values are relative, the bioptic (baseline) condition was fixed at 0 JOD for easier interpretation.

Results. Results in Figure 16-a show that our DiCE method produces images of higher perceived contrast compared to their standard/dioptic counterparts, demonstrating that the contrast fusion model is effective in complex images. The BTMO results are mixed, sometimes producing images of higher, but sometimes also of lower contrast compared to the standard/dioptic condition and DiCE. It is difficult to compare DiCE and BTMO techniques in terms of contrast enhancement, as each technique can produce images of even higher contrast if the binocular rivalry metric is relaxed. This, however, will result in images that are uncomfortable to view. The main strengths of DiCE over BTMO is that the enhancement is consistent across the images, demonstrating that the direct manipulation of contrast in DiCE offers better control over resulting images than the optimization used in the BTMO method.

The preference results, shown in Figure 16-b, are less conclusive as large subjective variations made most differences statistically insignificant. For the DiCE method, we could measure the preference difference only for the 2 out of 19 images. These differences could still be accidental as the test does not correct for multiple

[1] pwcmp — https://github.com/mantiuk/pwcmp

comparisons. For 8 out of 10 comparisons that are statistically significant, the BTMO method produced less preferred results than standard (dioptic) presentation and only in two cases the preference was higher. This is in contrast to findings from [Zhang et al. 2018], where the authors showed a strong preference for BTMO over standard presentation. We can only speculate that the effect could be due to the training of the participants; in our experiments, the participants with more exposure to dichoptic images also indicated stronger preference for them. This could be compared to the experience of wearing new glasses, when it takes some time to get fully comfortable and used to the new correction. This result could be also explained by the broad meaning of the "preference" criterion, which could combine many factors, such as comfort, familiarity, visual quality, wow-effect, etc. The results suggest that single-dimensional "preference" may not be the best measure for the dichoptic contrast enhancement techniques.

Figure 15 shows an example of two images produced by each method: the one for which BTMO produces a higher contrast image (Poker scene) and the one for which DiCE produces a higher contrast image (McKees Pub).

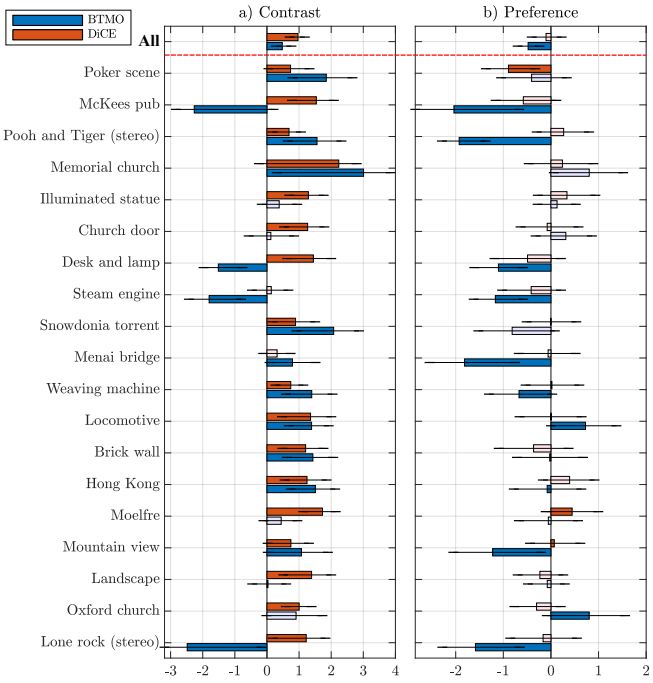

Fig. 16. The results of the validation experiment, comparing perceived contrast (a) and preference (b). The results are reported for each scene and for the aggregated results across all the scenes. The bars indicate the quality improvement relative to the standard presentation (no dichoptic enhancement) in JOD units (the higher the better). +1 JOD in that scale means that 75% of observers select the given condition over the standard presentation. The negative values mean that the standard condition is selected more often. The grayed bars indicate that we have no statistical evidence that a given condition is different (with respect to contrast or preference) from the standard presentation. The statistical test does not include the correction for multiple comparisons.

7.2 Experiment 3: Validation with VR

Experiment 2 was performed in a stereoscope, which provides high resolution and image quality, but it is less suitable for testing real-time rendering. Therefore, in the final experiment, we compare DiCE with standard presentation in VR environments. This experiment is also more relevant for the application of our method in real-time rendering. Note that we could not include BTMO in this experiment as that method is unsuitable for real-time rendering of 3D environments with 6DoF free viewing.

Apparatus and Participants. The VR environments were presented on an HTC Vive VR headset. Ten volunteers participated (2 female, mean age 25.8, SD 3.2 years).

Stimuli and Procedure. The stimuli consisted of 3 VR scenes shown in Figure 17, each seen from three different view points. The participants could freely look around the scenes while seated on a swivel chair. To switch between DiCE and standard presentation, the participants pressed the trigger on the Vive controller. We used l/h ratio of 0.55 for the DiCE method for similar reasons as in Experiment 2: to avoid overly conservative threshold adjustment and to make the method more different from the standard presentation. For each stimuli, they were asked three questions: "which scene appears to be higher in contrast?", "which scene appears to have more depth?" and "in which scene the materials and textures look more realistic?" (exact wording on the briefing form). The questions were motivated by our own observation that DiCE-enhanced images have different quality and appear more three-dimensional. We did not ask about their preference as the question did not give conclusive answers in Experiment 2. Before the experiment, each participant read and signed the briefing and consent forms. As the part of a training session, each participant was presented with three pairs of images with examples of low/high contrast, three-dimensional/flat shading, and natural/unnatural looking textures. We show the images for this training session in the supplementary materials. None of the participants reported symptoms of VR sickness after 10-15 minute session (no formal questionnaire was used).

Results. The results of Experiment 3 are presented in Figure 18 as percentages of participants who voted for DiCE when asked each of the three questions. It shows that our DiCE method produces higher contrast perception than standard presentation for all VR environments. The results also confirmed that the observers could see more depth with the DiCE enhancement. The effect can have a number of explanations. Ichihara et al.[2007] showed that increased contrast can give an impression of depth. Binocular luster may be causing lustrous features to pop out [Wolfe and Franzel 1988], giving the impression of false depth. Another possible explanation is that artificial disparity stemming from the different monocular images (luminance dichopticities) could give rise to a depth sensation [Wolfe 1986]. The results for realistic looking textures were less conclusive with only one environment, with the simplest textures and lowest complexity, showing moderate preference for DiCE.

8 DISCUSSION

The results of Experiment 2 and 3 confirmed that our DiCE technique can effectively enhance contrast not only for simplified stimuli,

Fig. 17. Three VR scenes in Experiment 3: Road, Rock, Woods (from left to right).

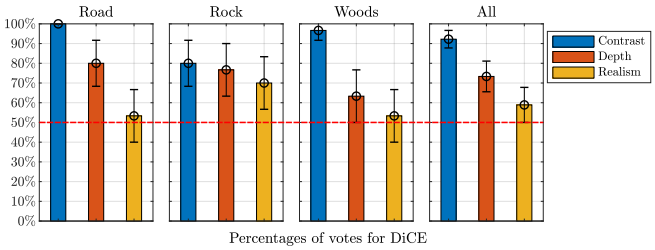

Fig. 18. The percentages of votes for DiCE when compared to the standard presentation. The results are reported for each VR environment and for the aggregated results across all the environments. The error bars denote the confidence intervals.

used in psychophysical models, but also for complex images. Experiment 3 delivered a more surprising finding, indicating that our technique can also improve impression of depth in images. This question emerged when we were inspecting the results of our method and noticed that they look different from typical monoscopic images because of an apparent impression of depth, even if such depth is false. We also noticed that materials change their appearance when processed with our technique. Glossy objects appear more shiny, giving them more realistic appearance. As we did not capture these effects in a formal experiment, we can only direct the readers to the supplementary materials to inspect the large collection of processed images. Full understanding of appearance changes caused by dichoptic presentation would require further research.

Contrary to previous BTMO techniques [Yang et al. 2012; Zhang et al. 2018] that venture to make images look different to the eyes but may or may not result in contrast enhancement, our method always enhances contrast in a principled manner. BTMO methods require very expensive optimization; for the 800×600 image the authors report 22.24 seconds per single iteration for the 2012 technique and 2.36 seconds for the 2018 technique. Our technique is given almost for free as most rendering pipelines include a tone curve, which can be customized per eye using our interleaved tone curves. Binocular rivalry can be controlled without the need for complex predictors, by simply changing a single parameter. We test our technique on both monoscopic and stereoscopic images, the latter being more relevant to the intended application.

One aspect of the binocular fusion that our method does not directly address is the ocular dominance of the user. The visual system has a preference for one of the eyes, especially in the presence of strong rivalry. Since our method attempts to reduce rivalry, ocular dominance is less relevant than image contrast. Legge & Rubin [1981] and Kingdom & Libenson [2015] showed that the eye receiving the higher contrast image dictates whether it contributes more to the fused image and the effect of eye dominance is not clearly visible in their data. However, we noted in our experiments that swapping images between the eyes results in changes in contrast perception. Our initial experiment on this issue (Experiment 4 in the supplementary) does not show conclusive results, but indicates that there may be an effect of eye dominance or other factors for some individuals, and thus that the fused contrast may not always be exactly as expected.

The main limitation of our technique is the inherent trade-off between contrast enhancement and binocular rivalry. Stronger levels of enhancement result in more rivalry, which is perfectly acceptable for some observers but not for others. This was evidenced in the preference results of Experiment 2, where the answers were mixed even though most observers reported seeing higher contrast. Clearly, more factors than perceived contrast contributed to the preference judgments. We suspect that the nature of the dichoptic enhancement requires some period of "wearing-in", similarly to getting used to a new pair of glasses. We did not observe any symptoms of VR sickness but such symptoms can be only revealed in a longer, purpose-designed experiment that has a control condition.

9 CONCLUSION

We propose a contrast enhancement technique for stereoscopic presentation, which is derived in a principled manner from a contrast fusion model. The main challenge of our approach is striking the right balance between contrast enhancement and visual discomfort caused by binocular rivalry. To address this challenge, we conducted a psychophysical experiment to test how content, observer, and tone curve parameters can influence binocular rivalry stemming from the dichoptic presentation. We found that the ratio of tone curve slopes can predict binocular rivalry letting us easily control the shape of the dichoptic tone curves. We validate the effectiveness of our technique in the evaluation study, in which we compare our technique with standard/dioptic presentation and previous techniques, for both monoscopic and stereoscopic images. We observed marked visual improvement in each case, particularly an unforeseen improvement in perceived depth. In addition, glossy objects show increased shininess and are thus perceived as more realistic. The technique has very low computational cost and only requires applying a separate tone curve for each eye. The single parameter of the

curve generation may be needed to be adjusted per observer but it is content independent so it does not require any analysis of the input images content, which is a costly operation in real-time rendering. As tone mapping is usually a part of the rendering pipeline, our technique can be easily combined with existing VR/AR rendering at almost no cost.

ACKNOWLEDGMENTS

We would like to thank Kuba Maruszczyk, Jonas Fiala, Minjung Kim, and all anonymous participants for their help with this project, as well as the authors of the Binocular Tone Mapping papers for processing the images for our experiments. We would also like to thank the contributors to the Unity AssetStore for making their scenes freely available: The Tales Factory for Photoscanned MoutainsRocks PBR, Sandro T for Flooded Grounds, and Shapes for Nature Starter Kit 2. This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement N° 765911 (RealVision), under the European Research Council (ERC) Consolidator Grant agreement N° 725253 (EyeCode) and Advanced Grant agreement N° 788065 (FUNGRAPH).

REFERENCES

Randolph Blake. 1989. A neural theory of binocular rivalry. *Psychological review* 96, 1 (1989), 145.

M. W. Cannon. 1985. Perceived contrast in the fovea and periphery. *Journal of the Optical Society of America. A, Optics and image science* 2, 10 (oct 1985), 1760–8. http://www.ncbi.nlm.nih.gov/pubmed/4056950

Felice A Dunn and Fred Rieke. 2008. Single-photon absorptions evoke synaptic depression in the retina to extend the operational range of rod vision. *Neuron* 57, 6 (mar 2008), 894–904. https://doi.org/10.1016/j.neuron.2008.01.031

Frédo Durand and Julie Dorsey. 2002. Fast bilateral filtering for the display of high-dynamic-range images. *ACM Transactions on Graphics* 21, 3 (jul 2002), 257–266. https://doi.org/10.1145/566654.566574

Gabriel Eilertsen, Rafał K. Mantiuk, and Jonas Unger. 2015. Real-time noise-aware tone mapping. *ACM Transactions on Graphics* 34, 6 (oct 2015), 1–15. https://doi.org/10.1145/2816795.2818092

G. Eilertsen, R. K. Mantiuk, and J. Unger. 2017. A comparative review of tone-mapping algorithms for high dynamic range video. *Computer Graphics Forum* 36, 2 (may 2017), 565–592. https://doi.org/10.1111/cgf.13148

Gabriel Eilertsen, Robert Wanat, Rafał K Mantiuk, and Jonas Unger. 2013. Evaluation of Tone Mapping Operators for HDR-Video. *Computer Graphics Forum* 32, 7 (oct 2013), 275–284. https://doi.org/10.1111/cgf.12235

Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski. 2008. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM SIGGRAPH 2008*. ACM, 1–10. http://portal.acm.org/citation.cfm?id=1399504.1360666

Philippe Fuchs. 2017. *Virtual reality headsets-a theoretical and pragmatic approach*. CRC Press.

M A Georgeson and G D Sullivan. 1975. Contrast constancy: deblurring in human vision by spatial frequency channels. *J. Physiol.* 252, 3 (nov 1975), 627–656.

Shigeru Ichihara, Norimichi Kitagawa, and Hiromi Akutsu. 2007. Contrast and Depth Perception: Effects of Texture Contrast and Area Contrast. *Perception* 36, 5 (2007), 686–695. https://doi.org/10.1068/p5696 arXiv:https://doi.org/10.1068/p5696 PMID: 17624115.

Fred Kingdom and Bernard Moulden. 1988. Border effects on brightness: A review of findings, models and issues. *Spatial Vision* 3, 4 (jan 1988), 225–262. https://doi.org/10.1163/156856888X00140

Frederick A. A. Kingdom and Lauren Libenson. 2015. Dichoptic color saturation mixture: Binocular luminance contrast promotes perceptual averaging. *Journal of Vision* 15, 5 (apr 2015), 2. https://doi.org/10.1167/15.5.2

Victor Klymenko, Robert W Verona, Howard H Beasley, and John S Martin. 1994. Convergent and divergent viewing affect luning, visual thresholds, and field-of-view fragmentation in partial binocular overlap helmet-mounted displays. In *Helmet-and Head-Mounted Displays and Symbology Design Requirements*, Vol. 2218. International Society for Optics and Photonics, 82–97.

J.J. Kulikowski. 1976. Effective contrast constancy and linearity of contrast sensation. *Vision Research* 16, 12 (jan 1976), 1419–1431. https://doi.org/10.1016/0042-6989(76)90161-9

G E Legge and J M Foley. 1980. Contrast masking in human vision. *Journal of the Optical Society of America* 70, 12 (dec 1980), 1458–71.

Gordon E Legge and Gary S Rubin. 1981. Binocular interactions in suprathreshold contrast perception. *Attention, Perception, & Psychophysics* 30, 1 (1981), 49–61.

Willem JM Levelt. 1965. Binocular brightness averaging and contour information. *British journal of psychology* 56, 1 (1965), 1–13.

R. Mantiuk, S. Daly, and L. Kerofsky. 2008. Display adaptive tone mapping. *ACM Transactions on Graphics* 27, 3 (2008), 68. https://doi.org/10.1145/1360612.1360667

Rafał K. Mantiuk, Karol Myszkowski, and Hans-peter Seidel. 2015. High Dynamic Range Imaging. In *Wiley Encyclopedia of Electrical and Electronics Engineering*. John Wiley & Sons, Inc., Hoboken, NJ, USA, 1–42. https://doi.org/10.1002/047134608X.W8265

Stephen E Palmer. 1999. *Vision science: Photons to phenomenology*. MIT press.

Sylvain Paris, Samuel W. Hasinoff, and Jan Kautz. 2011. Local Laplacian filters. *ACM Transactions on Graphics* 30, 4 (jul 2011), 1. https://doi.org/10.1145/2010324.1964963

Robert Patterson, Marc Winterbottom, Byron Pierce, and Robert Fox. 2007. Binocular rivalry and head-worn displays. *Human factors* 49, 6 (2007), 1083–1096.

Maria Perez-Ortiz and Rafal K. Mantiuk. 2017. A practical guide and software for analysing pairwise comparison experiments. *arXiv preprint* (dec 2017). arXiv:1712.03686 http://arxiv.org/abs/1712.03686

Erik Reinhard, Michael Stark, Peter Shirley, and James Ferwerda. 2002. Photographic tone reproduction for digital images. *ACM Transactions on Graphics* 21, 3 (jul 2002), 267. https://doi.org/10.1145/566654.566575

Peter Vangorp, Rafat K Mantiuk, Bartosz Bazyluk, Karol Myszkowski, Radosław Mantiuk, Simon J Watt, and Hans-Peter Seidel. 2014. Depth from HDR: depth induction or increased realism?. In *ACM Symposium on Applied Perception - SAP '14*. ACM Press, New York, New York, USA, 71–78. https://doi.org/10.1145/2628257.2628258

Gregory Ward-Larson, Holly Rushmeier, and Christine Piatko. 1997. A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes. *IEEE Transactions on Visualization and Computer Graphics* 3, 4 (1997), 291–306. https://doi.org/10.1109/2945.646233

Jeremy M Wolfe. 1986. Stereopsis and binocular rivalry. *Psychological review* 93, 3 (1986), 269.

Jeremy M. Wolfe and Susan L. Franzel. 1988. Binocularity and visual search. *Perception & Psychophysics* 44, 1 (01 Jan 1988), 81–93. https://doi.org/10.3758/BF03207480

Xuan Yang, Linling Zhang, Tien-Tsin Wong, and Pheng-Ann Heng. 2012. Binocular tone mapping. *ACM Transactions on Graphics (TOG)* 31, 4 (2012), 93.

Zhuming Zhang, Chu Han, Shengfeng He, Xueting Liu, Haichao Zhu, Xinghong Hu, and Tien-Tsin Wong. 2019. Deep binocular tone mapping. *The Visual Computer* 35, 6 (01 Jun 2019), 997–1011. https://doi.org/10.1007/s00371-019-01669-8

Zhuming Zhang, Xinghong Hu, Xueting Liu, and Tien-Tsin Wong. 2018. Binocular Tone Mapping with Improved Overall Contrast and Local Details. *Comput. Graph. Forum* 37, 7 (2018), 433–442. https://doi.org/10.1111/cgf.13580